

Research Article

Robust Hotelling T^2 Control Chart with Consistent Minimum Vector Variance

Hazlina Ali, Sharifah Soaad Syed Yahaya, and Zurni Omar

College of Art and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Correspondence should be addressed to Hazlina Ali; hazlina@uum.edu.my

Received 25 February 2013; Revised 21 August 2013; Accepted 25 August 2013

Academic Editor: A. Elías-Zúñiga

Copyright © 2013 Hazlina Ali et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, an alternative robust control chart based on a new robust estimator known as minimum vector variance (MVV) estimator, T^2_{MVV} , was introduced in Phase II. T^2_{MVV} was able to detect out-of-control signal and simultaneously control false alarm rate even as the dimension increased. However, the estimated UCLs of T^2_{MVV} are large as compared to the traditional chart. In this study, we improved the MVV estimators in terms of consistency and bias. The result showed great improvement in the control limit values while maintaining its good performance in terms of false alarm and probability of detection.

1. Introduction

Hotelling T^2 statistic was the first statistic known to be used in multivariate control chart. The control chart is referred to as Hotelling T^2 control chart. This statistic is used to measure the significance of the shifted distance from the out-of-control mean vector, μ_s , to the nominal mean vector, μ_0 , with the assumption that the covariance matrix remains constant at Σ_0 . The purpose of the control chart is to monitor the stability of a multivariate process in Phase I and II. Analysis in Phase I seeks to identify a stable historical data set (HDS). From this dataset, the in-control mean vector and the in-control variance-covariance matrix are estimated and later will be used in the Phase II analysis. A successful process monitoring in Phase II totally depends on the estimates of the parameters obtained from a stable HDS. However, the estimators are easily affected by unstable process, that is, multivariate outliers. The existence of outliers can violate the normality assumption. This violation may lead to the inflation of control limits and reduction of the probability of detection in Phase I, which consequently will cause the level of false alarm to be distorted, and the power to detect changes will be reduced in Phase II process [1]. False alarm rate is the probability of out-of-control signal when a process is in control. The value becomes large if the process is unstable due to the increase in variability. Inflated false alarm rate can lead to unnecessary process adjustments and loss of

confidence in the control chart as a monitoring tool [2]. Hence, a method which can control the false alarm rate to the desired (nominal) level is necessary.

However, the traditional Hotelling T^2 control chart is only effective in eliminating extreme outliers in small sample sizes, but it fails to detect moderate outliers particularly when the number of variables is large [3–5]. To overcome the problem, alternative estimation methods have been proposed in the literature. One of the approaches is to calculate the T^2 statistic based on successive differences variance-covariance matrix estimator [6–9]. Though this approach is effective in detecting shifts in the mean vector, it fails to detect other outliers as shown in Vargas [3]. Another approach is to use robust estimators in place of the classical estimators (\bar{x} and S). Robust estimators are known to be more effective in detecting the deviation of data, or outliers as compared to the classical estimators [10]. A wide range of robust estimators of multivariate location and scatter is available; see [11, 12] for a review. However, MCD estimator is more attractive than others because it has good theoretical properties with affine equivariance, high breakdown value, bounded influence function, and better convergence rate [13, 14]. The study on the significant role of MCD estimators in the construction of robust Hotelling T^2 chart can be easily found in the literature. Vargas [3] and Jensen et al. [4] introduced robust control chart based on MCD estimator for multivariate individual observations. They identified and removed the

outliers in Phase I analysis by using robust estimator and then calculated the classical estimator using the remaining good data points for Phase II analysis. They noticed some drawbacks when MCD was used in Phase I. Hotelling's T^2 issued from MCD needed a larger sample size if large number of outliers was suspected to ensure that MCD estimator did not breakdown and lose its ability especially when monitoring with more quality characteristics (p). To abate the problems, Chenouri et al. [5] proposed robust Hotelling T^2 chart based on reweighted MCD estimator. Besides possessing the nice properties of MCD estimator, the estimator was not unduly influenced by outliers and was more efficient than MCD. However, their approach was different from Vargas [3] and Jensen et al. [4] whereby they used RMCD estimator in place of classical estimators in constructing Hotelling T^2 chart for Phase II data directly. Using the same approach as Chenouri et al. [5], Alfaro and Ortega [15] conducted a comparison study on the performance of Hotelling T^2 control chart in Phase II process based on MCD, MVE, reweighted MCD, and trimmed estimator. Their finding showed a conflict between the percentage of outliers detection and the ability of the robust control charts in controlling the overall false alarm rate under certain conditions. To alleviate this conflict, Yahaya et al. [16] introduced the MVV estimator in Hotelling T^2 chart (T^2_{MVV}) in Phase II. In general, the result showed that T^2_{MVV} chart was able to increase the detection of out-of-control signals and simultaneously control false alarm rates even with large number of quality characteristics. In contrast, the MCD charts performed well in detecting out-of-control signals but failed in controlling false alarm rates. The traditional chart, however, was able to control false alarm rates but not effective in detecting out-of-control signals. Despite the good performance of T^2_{MVV} , the estimated UCLs for Hotelling T^2 chart issued from MVV estimators were large as compared to the traditional and MCD charts. Thus, this study attempts to improve the MVV estimators in achieving the desired UCLs by making the estimators consistent at normal model. Since in practice we always deal with finite samples, therefore the issue of bias in a finite sample is also considered in this study. The advantage of having unbiased estimator for a finite sample is that this estimator remains unbiased even though the sample size becomes larger [17]. With respect to the latter issue, this paper will also seek to improve the performance of MVV by making it unbiased for finite samples.

The organization of the remaining part of this paper is as follows. The formal definition of MVV estimator and the adjustment done on the MVV scatter estimator to ensure that it is consistent and unbiased will be discussed next, followed by the investigation on the improved MVV estimator through simulation study. The discussion continues with the computation of control limits for the traditional, MCD, and the improved Hotelling T^2_{MVV} charts, and the improvement of the proposed chart is revealed in this section. A real data analysis from aircraft industry is presented to illustrate the applicability of the proposed charts before arriving to the conclusion in the last section.

2. Minimum Vector Variance (MVV) Estimator

Herwindiati et al. [18] had proved that MVV estimators possess three major properties of a good robust estimator, that is, high breakdown point, affine equivariance, and computational efficiency. The main method used in the estimation of MVV is the Mahalanobis squared distances (MSDs). Let $X = \{x_1, x_2, \dots, x_n\}$ be a data set of p -variate observations. Denote the MVV estimators for the location parameter and scatter by M_{MVV} and S_{MVV} respectively. Now let $H \subseteq X$, the m_{MVV} and S_{MVV} are determined based on the set H consisting of $h = \lfloor (n + p + 1)/2 \rfloor$ data such that S_{MVV} has minimum trace of S^2_{MVV} , denoted as $\text{Tr}(S^2_{MVV})$, among all possible sets of h data. To compute the estimates of MVV, we used the MVV algorithm proposed in Yahaya et al. [16]. The location and scatter estimators are defined as

$$m_{MVV} = \frac{1}{h} \sum_{i=1}^h x_i, \quad (1)$$

$$S_{MVV} = \frac{1}{h} \sum_{i=1}^h (x_i - M_{MVV})(x_i - M_{MVV})^t. \quad (2)$$

2.1. Consistency Factor. The aim of Hotelling T^2 chart in Phase I is to estimate the in-control parameters of location, μ and scatter, Σ . The usual estimators for these parameters are the normal maximum likelihood estimators (MLE). The estimation of parameters is based on the data set $x = \{x_1, x_2, \dots, x_n\}$ from multivariate normal distribution with density

$$f_{\mu, \Sigma}(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-(x-\mu)^t \Sigma^{-1} (x-\mu)/2}, \quad (3)$$

with $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{Z}^+$. However, the distribution of (3) is only an approximation because a portion of the data may be contaminated by outliers [19]. With the existence of outliers, MLE which are known to be sensitive to outliers will not be able to precisely estimate the parameters. To address this problem, we propose MVV estimators, that is, robust estimators with highest breakdown point (50%) proposed by Herwindianti [20] to replace the MLE. We compute the MVV estimators in Phase I data sets, with location and scatter estimators as defined in (1) and (2), respectively. The MVV estimator has a fixed integer h such that

$$\left\lfloor \frac{n + p + 1}{2} \right\rfloor \leq h < n. \quad (4)$$

The preferred choice of h for outlier detection is its lower bound, which yields the breakdown value, $BP = (n - 2(p - 1))/2n$. Let m_{MVV} and S_{MVV} be the mean and the scatter matrix calculated from the h observations out of x_i , whose classical scatter matrix has the lowest vector variance resulting from h smallest MSD. The S_{MVV} is a scatter $p \times p$ matrix which is positive definite, symmetric (PDS), and affine equivariant [20]. However, this estimator is not consistent under normal model. Robust scatter estimator is

TABLE 1: Values of the $\vartheta_{p,n}^\alpha$ for $c(h)S_{\text{MVV}}$.

$p = 2$		$p = 5$		$p = 10$		$p = 15$		$p = 20$	
n	$\vartheta_{p,n}^\alpha$	n	$\vartheta_{p,n}^\alpha$	n	$\vartheta_{p,n}^\alpha$	n	$\vartheta_{p,n}^\alpha$	n	$\vartheta_{p,n}^\alpha$
10	5.8276	30	2.9304	50	2.3127	80	1.9801	100	1.8503
25	5.4200	50	2.9598	100	2.2098	100	1.9528	200	1.7762
50	5.7715	100	2.9309	200	2.1524	200	1.8937	300	1.7490
100	5.7206	200	2.8912	500	2.1045	500	1.8476	500	1.7180
200	5.6679	500	2.8579						
500	5.5842								

typically calibrated to be consistent for normal model. Known as Fisher consistency, this is a standard concept in robust statistics which denotes that the functionals evaluated at the model distribution return the true parameter value, Σ [19]. In order to achieve consistency under the normal model, S_{MVV} (in (2)) is multiplied by a consistency factor, $c(h)$, as follows:

$$c(h)S_{\text{MVV}} = \frac{c(h)}{h} \sum_{i=1}^h (x_i - m_{\text{MVV}})(x_i - m_{\text{MVV}})^t. \quad (5)$$

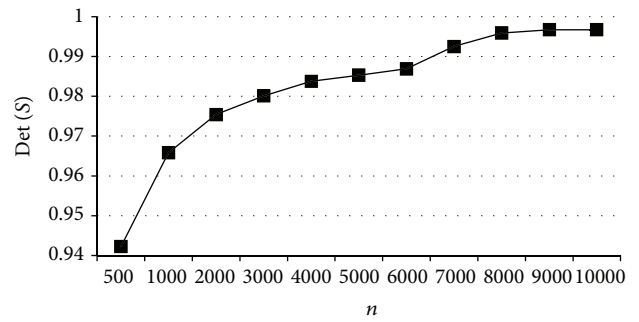
The approximation of consistency factor can be obtained from elliptical truncation in the multivariate normal distribution based on squared distance. If $x_i \sim N(\mu, \Sigma)$, $c(h)$ is defined as

$$c(h) = \frac{h/n}{P(\chi_{p+2}^2 < \chi_{p,h/n}^2)}, \quad (6)$$

where $\chi_{p,h/n}^2$ is the h/n -quantile of χ_p^2 distribution. This formula is derived by Butler et al. [13] and further discussed in Croux and Haesbroeck [14] based on the functional form of the MCD estimator. Since MVV have the same functional form with the MCD estimator, we used (6) as the consistency factor for S_{MVV} . Albeit guaranteed consistency under normality distribution, Pison et al. [17] cautioned that MCD estimators were biased for small sample sizes. Thus, the consistency factor in (6) only might not be sufficient to make MVV estimator unbiased for small sample sizes. For that reason, we also include the computation of correction factor at any sample size n and dimension p .

2.2. Correction Factor. A simulation study on the effect of correction factor on the MVV estimator is carried out for several sample sizes n and dimension $p = 2, 5, 10, 15$, and 20. We generated data sets $X^{(j)} \in \mathbb{R}^{n \times p}$ from standard multivariate normal distribution. For each data set $X^{(j)}$, $j = 1, \dots, m$ we then determine the $c(h)S_{\text{MVV}}^{(j)}$ in (5). If the estimator is unbiased, $E[c(h)S_{\text{MVV}}] = I_p$, therefore the p th root of the determinant of $c(h)S_{\text{MVV}}$ equals 1. Given the mean of the p th root of the determinant as $\text{mean}(|c(h)S_{\text{MVV}}|) = (1/m) \sum_{j=1}^m (|c(h)S_{\text{MVV}}^{(j)}|)^{1/p}$. To determine the correction factor, we performed $m = 1000$ simulations for different sample sizes n and dimensions p , with $\alpha = 0.05$ such that

$$\vartheta_{p,n}^\alpha = \frac{1}{\text{mean}(|c(h)S_{\text{MVV}}|)}. \quad (7)$$

FIGURE 1: Determinant of $\vartheta_{p,n}^\alpha c(h)S_{\text{MVV}}$ when $n \rightarrow \infty$ and $p = 10$.

The computed values are displayed in Table 1. Then, using $\vartheta_{p,n}^\alpha$ in (7) as the correction factor for $c(h)S_{\text{MVV}}$, we obtain

$$\vartheta_{p,n}^\alpha c(h)S_{\text{MVV}} = \frac{\vartheta_{p,n}^\alpha c(h)}{h} \sum_{i=1}^h (x_i - m_{\text{MVV}})(x_i - m_{\text{MVV}})^t. \quad (8)$$

Since $\vartheta_{p,n}^\alpha c(h)S_{\text{MVV}}$ can be considered consistent and unbiased, the determinant of $\vartheta_{p,n}^\alpha c(h)S_{\text{MVV}}$ should approach 1.

3. Investigation through Simulation Experiment

Gather and Becker [21] have emphasized that robust estimators to be used in the method of outliers detection should have sufficient rate of convergence to some true underlying model parameter for consistency and unbiased. A sequence of asymptotically unbiased estimators for parameter θ is called consistent if $\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0$. To illustrate the analysis on the consistency of MVV estimator at multivariate normal, data are randomly generated from $N(0, I_p)$. An experiment is carried out for several values of sample sizes n until convergent for a fixed moderate dimension such that $p = 10$. Figure 1 shows the determinants of $\vartheta_{p,n}^\alpha c(h)S_{\text{MVV}}$ corresponding to the sample size, n . As the value of n increases, we can observe that the determinant approaches 1 which implies that the $\vartheta_{p,n}^\alpha c(h)S_{\text{MVV}}$ is consistent. Next, the investigation using simulation experiment continues to show that m_{MVV} and $\vartheta_{p,n}^\alpha c(h)S_{\text{MVV}}$ which replaced the MLE, μ and Σ , in Hotelling T^2 are consistent and

unbiased. The squared distances using any affine-equivariant robust location and scatter estimators which are consistent and unbiased under normal model are asymptotically χ^2 distributed [21]. Therefore, if m_{MVV} and $\vartheta_{p,n}^\alpha c(h) S_{\text{MVV}}$ are consistent and unbiased estimators for μ and Σ , then with observations x_i i.i.d in $\mathbb{R}^p \sim N_p(\mu, \Sigma)$, it follows that $d_i^2 = (x_i - m_{\text{MVV}}) \vartheta_{p,n}^\alpha c(h) S_{\text{MVV}}^{-1} (x_i - m_{\text{MVV}})^t$ is asymptotically χ_p^2 distributed. Since d_i^2 is similar to Hotelling T^2 , the asymptotic distribution of the improved Hotelling T_{MVV}^2 when $n \rightarrow \infty$ should follow χ_p^2 distribution if the estimators are consistent. If we consider a sample of p quality characteristics such that $x_i = \{x_{i1}, x_{i2}, \dots, x_{ip}\}$ where $i = 1, 2, \dots, n$ as a phase I data set, then the improved T_{MVV}^2 statistic for x_i can be constructed in the following manner:

$$T_{\text{MVV}(I)}^2(i) = (x_i - m_{\text{MVV}}) \vartheta_{p,n}^\alpha c(h) S_{\text{MVV}}^{-1} (x_i - m_{\text{MVV}})^t. \quad (9)$$

To check on the distributions of the improved T_{MVV}^2 , we employed the QQ plots and evaluated the goodness of fit on those plots based on the slope and the R -square of the straight line as shown in Table 2. The hypothetical distribution represents the χ_p^2 without error if all points are in a straight line with slope equals 1 and R -square also equals 1 [22]. Random data were generated from multivariate standard normal distribution $MVN(0, I_p)$. This study is carried out for the sample size of $n = 10,000$ with dimensions of $p = 2, 5, 10, 15$, and 20 . From this table we observe that the R -square values for all p 's are 0.999. With regard to the slopes, we can see a considerable difference in the values between the Hotelling's T^2 with original MVV ($T_{\text{MVV}(o)}^2$) and Hotelling's T^2 with improved MVV ($T_{\text{MVV}(I)}^2$) especially when $p = 2$. The slopes for $T_{\text{MVV}(I)}^2$ are consistent and approximately equal to 1 regardless of the dimensions (p). In contrast, the slopes for $T_{\text{MVV}(o)}^2$ are quite a distance away from 1 even though the pattern shows a declining in values towards 1 as p increases. We observe that the values for the two measurements (R^2 and slopes) are very close to the ideal value, which signify that the χ_p^2 distribution fits well with the simulated $T_{\text{MVV}(I)}^2$ values. The result implies that the constant $\vartheta_{p,n}^\alpha c(h)$ fulfills the condition of the multiplicative factors to make the S_{MVV} estimators consistent and unbiased for Σ .

4. MVV Hotelling T^2 Control Chart

Let $x_i = \{x_{i1}, x_{i2}, \dots, x_{ip}\}$ be the p -variate random sample of n observations of preliminary data set in Phase I. Calculate the M_{MVV} and $\vartheta_{p,n}^\alpha c(h) S_{\text{MVV}}$ estimators. Since the estimators are known to be free from outliers due to their estimation process, they could be readily used as in-control estimators in Phase II. By using these estimates, $T_{\text{MVV}(I)}^2$ statistic in (9) is computed for Phase II observation, $x_g = \{x_{n+1}, x_{n+2}, \dots\}$ where $x_g \notin x_i$.

4.1. Estimation of Control Limits. In this section, we present the control limit of the improved $T_{\text{MVV}(I)}^2$ control chart by

TABLE 2: The slope and R -square for $T_{\text{MVV}(o)}^2$ and $T_{\text{MVV}(I)}^2$.

$n = 10\,000$		$T_{\text{MVV}(o)}^2$	$T_{\text{MVV}(I)}^2$
$p = 2$	R^2	0.999	0.999
	slope	3.264	1.001
$p = 5$	R^2	0.999	0.999
	slope	1.952	1.020
$p = 10$	R^2	0.999	0.999
	slope	1.538	1.003
$p = 15$	R^2	0.999	0.999
	slope	1.404	1.001
$p = 20$	R^2	0.999	0.999
	slope	1.341	1.005

using simulated data with different combinations of sample sizes, n , and number of dimensions, p . The control limit of $T_{\text{MVV}(I)}^2$ chart is then compared with the control limit of $T_{\text{MVV}(o)}^2$ chart, robust Hotelling T^2 chart using MCD (T_{MCD}^2) and the traditional Hotelling T^2 charts. The application of robust estimators in place of the mean and covariance structure in traditional Phase II Hotelling T^2 statistic will cause the distributional properties of the traditional chart to change [9]. To demonstrate the performance of $T_{\text{MVV}(o)}^2$ and $T_{\text{MVV}(I)}^2$ control charts, we need to identify the distribution of each method in order to obtain appropriate control limits, that is, UCL. Since the exact distribution of $T_{\text{MVV}(o)}^2$ is unknown, we apply Monte Carlo method to estimate the quantiles of the $T_{\text{MVV}(o)}^2$ and $T_{\text{MVV}(I)}^2$, for several combinations of sample sizes and dimensions. In order to estimate the 95% quantile of $T_{\text{MVV}(I)}^2$ for a given Phase I of sample size n and dimension p , we generate $K = 5000$ samples of size n from a standard multivariate normal distribution, $MVN_p(0, I_p)$. For each data set of size n , we compute the MVV mean vector and the modified covariance matrix estimates, $M_{\text{MVV}}(k)$ and $\vartheta_{p,n}^\alpha c(h) S_{\text{MVV}}(k)$, respectively, from $k = 1, \dots, K$. In addition, for each data set, we randomly generate a new observation $x_{g,k}$ treated as a Phase II observation from $MVN_p(0, I_p)$ and calculate the corresponding $T_{\text{MVV}(I)}^2(g, k)$ values. The empirical distribution function of $T_{\text{MVV}(I)}^2(g)$ is based on the simulated values

$$T_{\text{MVV}(I)}^2(g, 1), T_{\text{MVV}(I)}^2(g, 2), \dots, T_{\text{MVV}(I)}^2(g, K). \quad (10)$$

We sort $T_{\text{MVV}(I)}^2(g, k)$ values in ascending order, and the UCL is the 95% quantile of the 5000 statistics. The results of the investigation are presented in Table 3. We observe that the estimated UCLs for $T_{\text{MVV}(o)}^2$ are large as compared to the traditional control charts T_S^2 , T_{MCD}^2 , and T_{RMCD}^2 . However, after making the MVV scatter estimator consistent and unbiased as shown in (8), the results improved immensely. As we can see here, the UCLs are closer to the traditional UCLs.

4.2. Real Data Analysis. The application of the improved method $T_{\text{MVV}(I)}^2$ on real data is illustrated using data furnished by Asian Composites Manufacturing Sdn. Bhd. (ACM) which involves in the production of advanced

TABLE 3: Control limits of the investigated control charts for various combinations of sample sizes and dimensions.

p	n	T^2	T^2_{MCD}	$T^2_{MVV(o)}$	$T^2_{MVV(I)}$
2	10	11.0360	30.6250	76.0122	19.7067
	25	7.4275	12.1798	32.4008	9.4443
	50	6.6447	8.2762	28.1107	8.0556
	100	6.3039	7.4463	24.6037	7.1969
	200	6.1443	6.4127	21.7088	6.4469
	500	6.0518	6.1558	20.4264	6.1788
5	30	15.6006	27.6404	41.9567	19.5315
	50	13.4506	18.3456	33.5214	15.9398
	100	12.1579	14.7736	28.4822	14.0082
	200	11.5915	12.5765	25.6204	12.9297
	500	11.2738	11.4941	22.5296	11.5868
10	50	25.9552	39.8024	62.9323	34.5417
	100	21.5264	26.0646	43.1889	25.5450
	200	19.7975	20.9145	34.8509	21.4812
	500	18.8777	19.5618	31.1418	19.8112
15	80	33.6517	42.7942	69.0937	42.6982
	100	31.5083	37.4367	61.0544	38.6310
	200	27.9034	29.4809	45.9981	30.6107
	500	26.0882	26.6016	39.7181	26.1456
20	100	42.5747	52.7273	83.5238	53.9627
	200	36.2033	38.8163	57.2303	39.3577
	300	34.4609	36.0595	52.3438	36.8230
	500	33.1766	33.6574	47.7808	34.4221

composite panels for the aircraft industry. ACM produces flat and contoured primary (Aileron Skins, Spoilers and Spars) and secondary (Flat Panels, Leading Edges and MISC: Components) structure composite bond assemblies and sub-assemblies for aerospace industries. The company provided us with data on spoilers as shown in Table 4. For the purpose of this study, a sample of 47 spoilers ($n = 47$) which consists of several features, namely, trim edge (X_1), trim edge spar (X_2), and drill hole (X_3) were furnished to us by the company. Out of the total, 21 spoilers were collected from 2009, while the rest were from 2010. Hence, we decided to use the 2009 spoilers as Phase I historical data and considered the spoilers from 2010 as future data in this study. Estimates for the location vector (\bar{X}) and scatter matrix (S) are presented in Table 5. In the last column of Table 5, we could clearly observe that the upper control limit (UCL) for $T^2_{MVV(I)}$ is the closest to the traditional Hotelling T^2 with values of 11.5513 and 11.035, respectively, whereas the other control charts produce large UCL values especially the original T^2_{MVV} ($T^2_{MVV(o)}$). When we compare the improved with the original T^2_{MVV} , we observe a large disparity between the two values such that $T^2_{MVV(o)} = 41.298$, and $T^2_{MVV(I)} = 11.5513$. The result indicates great improvement in the UCL values from $T^2_{MVV(o)}$ to $T^2_{MVV(I)}$. Table 6 identifies the out-of-control data (bold font) using the different T^2 statistics. Among the four statistics, $T^2_{MVV(o)}$, $T^2_{MVV(I)}$ and

TABLE 4: List of Phase I and Phase II real data.

No. of product	Phase I data			Phase II data		
	X_1	X_2	X_3	X_1	X_2	X_3
1	-0.0011	0.0003	0.0128	0.0041	0.0087	0.0129
2	0.0011	0.0021	0.0246	0.0047	0.0109	0.0124
3	0.0252	0.0308	0.0378	0.0031	0.0057	0.0096
4	-0.0017	0.0109	0.0177	0.0035	-0.0020	0.0101
5	-0.0005	-0.0010	0.0106	0.0040	-0.0028	0.0125
6	0.0016	-0.0059	0.0128	0.0031	0.0008	0.0061
7	0.0004	0.0001	0.0062	-0.0019	0.0101	0.0112
8	0.0078	0.0003	0.0159	0.0009	0.0039	0.0082
9	0.0076	0.0089	0.0097	-0.0052	0.0090	0.0203
10	0.0020	0.0005	0.0071	-0.0008	0.0110	0.0184
11	0.0108	0.0011	0.0092	-0.0021	0.0139	0.0170
12	0.0039	0.0034	0.0425	-0.0017	0.0092	0.0061
13	0.0060	-0.0033	0.0160	-0.0010	0.0133	0.0138
14	0.0066	0.0100	0.0056	-0.0030	0.0002	0.0053
15	0.0045	-0.0067	0.0147	0.0016	0.0134	0.0151
16	0.0110	-0.0207	0.0337	0.0027	0.0086	0.0070
17	0.0047	0.0059	0.0065	0.0004	0.0086	0.0087
18	0.0077	0.0003	0.0191	-0.0036	0.0136	0.0129
19	0.0015	0.0123	0.0124	-0.0028	0.0003	0.0078
20	0.0011	0.0038	0.0104	0.0120	0.0123	0.0768
21	0.0056	0.0065	0.0063	-0.0015	0.0004	0.0115
22				0.0009	0.0232	0.0202
23				-0.0035	0.0088	0.0107
24				0.0016	0.0061	0.0066
25				-0.0228	-0.0466	0.0231
26				0.0037	-0.0038	0.0147

T^2_{MCD} consider signal observations 20, 22, and 25 as out-of-control but T^2 fails to signal observation 22 and only consider, observations 20 and 25 as out-of-control. This is expected due to the low probability of detection in the traditional control chart [16]. For a clearer visualisation on the performance of the control charts in detecting out of control observations, graphical presentation of the corresponding control charts are put on view in Figure 2.

5. Conclusion

The UCL value for the Hotelling T^2 control chart using consistent and unbiased MVV estimators seemed to improve significantly from the Hotelling T^2 control chart based on the original MVV estimators. The improved control chart ($T^2_{MVV(I)}$) was put to test on real data. Even though the performance of the improved T^2_{MVV} control chart was on par with the original T^2_{MVV} chart, the improved estimators have successfully reduced the inflated UCL of the original T^2_{MVV} close to the UCL of the traditional Hotelling T^2 (T^2) control chart. However, when the improved control chart was compared with the traditional chart based on their

TABLE 5: Estimates of location vector, covariance matrix, and UCL.

Types of control chart	Location Vector (\bar{X})	Scatter Matrix (S)	Upper Control Limit (UCL)
T^2	$[0.00504 \ 0.00284 \ 0.01579]$	$\begin{bmatrix} 0.00004 & 0.00002 & 0.00003 \\ 0.00002 & 0.00009 & 0.00001 \\ 0.00003 & 0.00001 & 0.00011 \end{bmatrix}$	11.035
T^2_{MCD}	$[0.00414 \ 0.00207 \ 0.01096]$	$\begin{bmatrix} 0.00002 & 0.00000 & 0.00000 \\ 0.00002 & 0.00009 & -0.00002 \\ 0.00000 & -0.00002 & 0.00003 \end{bmatrix}$	21.946
$T^2_{MVV(o)}$	$[0.00336 \ 0.00354 \ 0.00913]$	$\begin{bmatrix} 0.00001 & 0.00001 & 0.00000 \\ 0.00001 & 0.00003 & 0.00000 \\ 0.00000 & 0.00000 & 0.00001 \end{bmatrix}$	41.298
$T^2_{MVV(l)}$	$[0.00336 \ 0.00354 \ 0.00913]$	$\begin{bmatrix} 0.00003 & 0.00002 & -0.00001 \\ 0.00002 & 0.00007 & -0.00001 \\ -0.00001 & -0.00001 & 0.00002 \end{bmatrix}$	11.5513

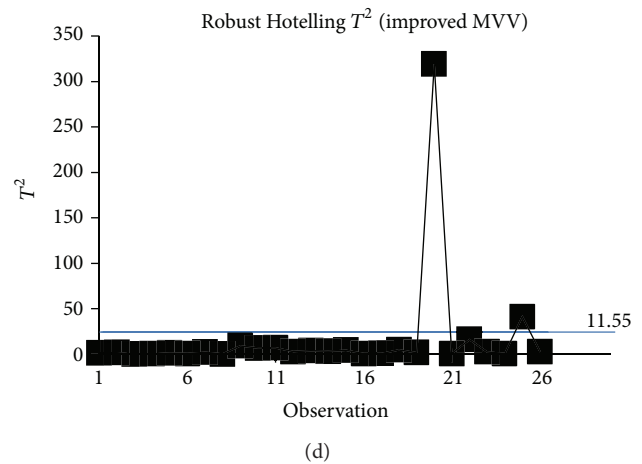
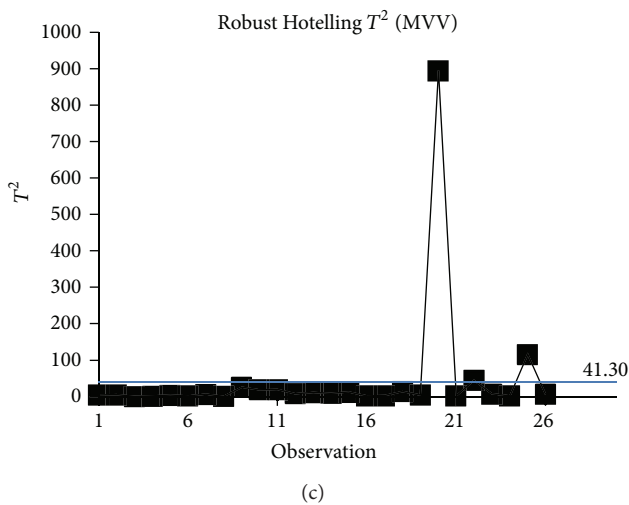
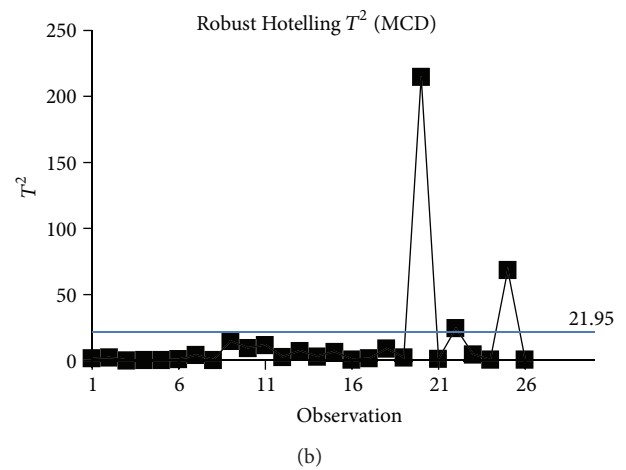
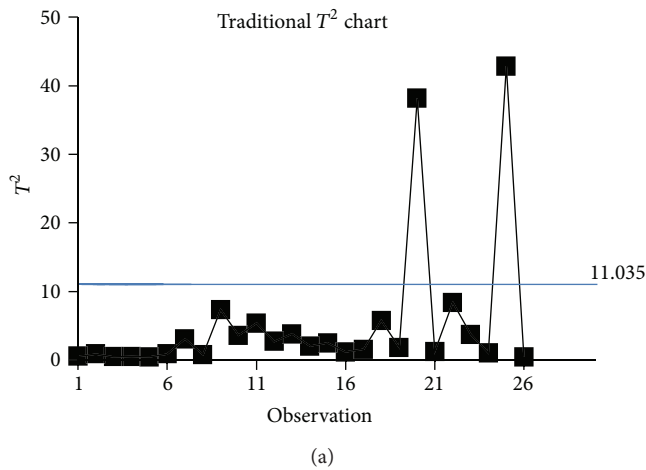
FIGURE 2: Hotelling T^2 Control Charts.

TABLE 6: The Hotelling T^2 values for the future (Phase II) data.

No.	T^2	T^2_{MCD}	$T^2_{MVV(o)}$	$T^2_{MVV(l)}$
1	0.55822	1.76591	4.39082	1.56608
2	0.90026	2.46944	5.16947	1.84380
3	0.49916	0.34367	0.29917	0.10671
4	0.54633	0.54563	1.50640	0.53729
5	0.45922	0.45797	3.78687	1.35067
6	0.90130	1.25274	2.24213	0.79971
7	3.09329	4.44043	6.53612	2.33125
8	0.80608	0.68370	1.05554	0.37648
9	7.36021	14.97663	26.04990	9.29127
10	3.61976	9.74168	19.17603	6.83955
11	5.38392	11.87166	19.63128	7.00193
12	2.73870	2.97882	8.13879	2.90288
13	3.80577	7.40398	11.38954	4.06233
14	2.05480	3.30863	9.14983	3.26349
15	2.50731	6.80538	12.38812	4.41850
16	1.19755	1.06789	2.05633	0.73344
17	1.57979	1.75966	2.87650	1.02597
18	5.79103	9.28168	13.92929	4.96819
19	1.83044	2.41775	4.87909	1.74023
20	38.13972	214.92329	894.51844	319.04971
21	1.26507	1.54862	2.06405	0.73619
22	8.41812	24.65515	45.24620	16.13805
23	3.75884	4.87934	7.53275	2.68672
24	1.06020	0.93200	2.23575	0.79743
25	42.84468	68.63065	116.02933	41.38442
26	0.48316	0.77959	7.32655	2.61318

almost equal UCLs, the finding showed that the improved control chart performed better in detecting out-of-control observations. With the good properties and performance, this improved MVV estimators that should be considered as alternative estimators to replace the usual mean and variance vector in the construction of the robust Hotelling T^2 control chart as well as other multivariate statistical procedures.

References

- [1] N. Ye, S. M. Emran, Q. Chen, and S. Vilbert, "Multivariate statistical analysis of audit trails for host-based intrusion detection," *IEEE Transactions on Computers*, vol. 51, no. 7, pp. 810–820, 2002.
- [2] Y. S. Chang and D. S. Bai, "A multivariate T^2 control chart for skewed populations using weighted standard deviations," *Quality and Reliability Engineering International*, vol. 20, no. 1, pp. 31–46, 2004.
- [3] J. A. Vargas, "Robust estimation in multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 35, no. 4, pp. 367–376, 2003.
- [4] W. A. Jensen, J. B. Birch, and W. H. Woodall, "High breakdown estimation methods for Phase I multivariate control charts," *Quality and Reliability Engineering International*, vol. 23, no. 5, pp. 615–629, 2007.
- [5] S. Chenouri, S. H. Steiner, and A. M. Variyath, "A multivariate robust control chart for individual observations," *Journal of Quality Technology*, vol. 41, no. 3, pp. 259–271, 2009.
- [6] D. S. Holmes and A. E. Mergen, "Improving the performance of the T^2 control chart," *Quality Engineering*, vol. 5, no. 4, pp. 619–625, 1993.
- [7] J. H. Sullivan and W. H. Woodall, "A comparison of multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 28, no. 4, pp. 398–408, 1996.
- [8] J. H. Sullivan and W. H. Woodall, "Adapting control charts for the preliminary analysis of multivariate observations," *Communications in Statistics B*, vol. 27, no. 4, pp. 953–979, 1998.
- [9] J. D. Williams, W. H. Woodall, J. B. Birch, and J. H. Sullivan, "Distribution of hotelling's T^2 statistic based on the successive differences estimator," *Journal of Quality Technology*, vol. 38, no. 3, pp. 217–229, 2006.
- [10] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, NY, USA, 1986.
- [11] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*, John Wiley & Sons, New York, NY, USA, 2006.
- [12] R. A. Maronna and R. H. Zamar, "Robust estimates of location and dispersion for high-dimensional datasets," *Technometrics*, vol. 44, no. 4, pp. 307–317, 2002.
- [13] R. W. Butler, P. L. Davies, and M. Jhun, "Asymptotics for the minimum covariance determinant estimator," *The Annals of Statistics*, vol. 21, no. 3, pp. 1385–1400, 1993.
- [14] C. Croux and G. Haesbroeck, "Influence function and efficiency of the minimum covariance determinant scatter matrix estimator," *Journal of Multivariate Analysis*, vol. 71, no. 2, pp. 161–190, 1999.
- [15] J. L. Alfaro and J. F. Ortega, "A comparison of robust alternatives to Hotelling's T^2 control chart," *Journal of Applied Statistics*, vol. 36, no. 12, pp. 1385–1396, 2009.
- [16] S. S. Yahaya, H. Ali, and Z. Omar, "An alternative hotelling T^2 control chart based on minimum vector variance (MVV)," *Modern Applied Science*, vol. 5, no. 4, pp. 132–151, 2011.
- [17] G. Pison, S. Van Aelst, and G. Willems, "Small sample corrections for LTS and MCD," *Metrika*, vol. 55, no. 1-2, pp. 111–123, 2002.
- [18] D. E. Herwindiati, M. A. Djauhari, and M. Mashuri, "Robust multivariate outlier labeling," *Communications in Statistics*, vol. 36, no. 6, pp. 1287–1294, 2007.
- [19] C. Croux and P. J. Rousseeuw, "A class of high-breakdown scale estimators based on subranges," *Communications in Statistics*, vol. 21, no. 7, pp. 1935–1951, 1992.
- [20] D. E. Herwindiati, *A new criterion in robust estimation for location and covariance matrix, and its application for outlier labeling [Ph.D. thesis]*, Institut Teknologi Bandung, Bandung, Indonesia, 2006.
- [21] U. Gather and C. Becker, "Convergence rates in multivariate robust outlier identification," in *Mathematics*, vol. 34, pp. 101–107, 1997.
- [22] H. Ali, M. A. Djauhari, and S. S. Yahaya, "On the distribution of FMCD-based robust mahalanobis distance," in *Proceedings of the 3rd International Conference on Mathematics and Statistics (ICoMS '08)*, InstitutPertanian Bogor, West Java, Indonesia, 2008, Paper no: 134-1506.

